creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

이학석사 학위논문

Multi-group analysis using generalized additive kernel canonical correlation analysis

일반화 가법 커널 정준상관분석을 이용한 다중그룹 분석

2019 년 6 월

서울대학교 대학원

통계학과

배 은 성

Multi-group analysis using generalized additive kernel canonical correlation analysis

일반화 가법 커널 정준상관분석을 이용한 다중그룹 분석

지도교수 임 채 영

이 논문을 이학석사 학위논문으로 제출함

2019 년 6 월

서울대학교 대학원

통계학과

배 은 성

배은성의 이학석사 학위논문을 인준함

2019 년 6 월

위 원 장 오 희 석 (인)

부위원장 임 채 영 (인)

위 원 임 요 한 (인)

Abstract

Multi-group analysis using generalized additive kernel canonical correlation analysis

Eunseong Bae
Department of Statistics
The Graduate School
Seoul National University

Multivariate analysis methods have been widely used and one of popular methods is canonical correlation analysis (CCA). Despite several advantages of CCA, it has some limitations; restricted to linear relationship and two groups. To overcome such limitation, modified version of CCA have been proposed by several researchers, like kernel CCA and generalized CCA. In this paper, we propose an extension of CCA that allows multi-group and nonlinear relationship in additive fashion. We call our approach Generalized Additive Kernel Canonical Analysis (GAKCCA). In addition to exploring multi-group relationship with nonlinear extension, GAKCCA can reveal contribution of variables in each group; which enables in-depth structural analysis. Simulation study shows that GAKCCA can distinguish a relationship between groups and whether they are correlated or not.

Keywords: Multivariate analysis, Generalized Additive Kernel Canonical Analysis, Multiblock data analysis

Student Number: 2017-25680

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Multivariate analysis is one of statistical methods that considers several variables simultaneously. Compared with univariate analysis, which focus on the influence of one variable only, multivariate analysis takes into account not only the effect of each variable but also interaction between variables. Thus, multivariate analysis gets popular as researchers face to more complex data. A number of statistical methods concerning multivariate analysis have been developed and widely used. For instance, principle component analysis (PCA), first proposed by Pearson (1901) [15] is a method to compress data in a high dimensional space into the low dimensional space. Factor analysis extracts underlying, but unobservable, random quantities by assuming variables are expressed with those random quantities ([9]).

One of the popular multivariate analysis is canonical correlation analysis (CCA). CCA, proposed by Hotelling (1936) [8], explores a relationship between two multivariate groups. CCA finds linear combinations of each group that maximize their Pearson correlation coefficient. In this way, CCA can serve as a dimension reduction method by analyzing linear combinations of each multidimensional variable and this advantage makes CCA widely used in many scientific fields that mostly deal with high dimensional data such as psychology, neuroscience, medical science and image recognition ([16, 22, 14, 24]), etc.

Despite of this strength, CCA has some limitations. First, CCA is restricted to linear relationship only so that CCA could result in bad performance if two variables are linked with a non-linear relation. This limitation is inherited from the characteristics of Pearson correlation. If two random variables X and Y are related with the equation $X^2 + Y^2 = 1$, then the Pearson correlation of X and Y results in $\mathrm{Corr}(X, Y) = 0$, although two random variables are related. To overcome the linearity constraint of classical CCA, Bach and Jordan (2002) [3] proposed Kernel canonical correlation analysis (KCCA), which applies a kernel method to the CCA problem. Unlike CCA, KCCA is a method of finding nonlinear relationship between two groups. Kernelization allows practical parametric nonlinear extension of the CCA method. KCCA has been used in some scientific fields that need to find nonlinear relationship beyond linear one such as speech communication science, genetics and pattern recognition ([2, 12, 25]).

Another limitation of classical CCA is that it is applicable only to two groups. Often, many scientific experiments yield results that can be grouped into more than two groups. Pair-wise application of CCA into the groups more than two may ignore the connection and non-connection within them. Multi-group version of CCA was introduced by Kettenring (1971) [10], named generalized canonical correlation analysis (GCCA or MCCA). GCCA finds linear combinations of each group that optimize certain criterion, like the sum of covariances. Tenenhaus *et al.* (2015) [20] proposed kernelized version of GCCA : kernel generalized canonical correlation analysis (KGCCA). This method is an extension of CCA by combining nonlinearity and multi-group analysis. In spite of fully flexible extension, kernelization of all variables together in each group cannot provide the structural analysis of CCA. For instance, KGCCA cannot reveal the contribution of X_{11} in $\boldsymbol{X}_1 = (X_{11}, \cdots, X_{1p_1})$ in a relation to another group $\boldsymbol{X}_2 = (X_{21}, \cdots, X_{2p_2})$. Balakrishnan *et al.* (2012) [5] considered

an additive model by restricting possible non-linear functions to the class of additive models; $f(x_1, \cdots, x_p) = \sum_{i=1}^{p} f_i(x_i)$. This modification enables KCCA to analyze the contribution of each variable. However, it is still restricted to two groups.

In this paper, we apply an additivity idea to KGCCA so that the proposed method allows more than two groups with nonlinear structure in an additive way. We call our proposed approach *generalized additive kernel canonical correlation analysis* (GAKCCA). We expect the proposed approach has a better interpretability than KCCA or KGCCA and it can be applied to multi-group data.

The organization of the paper is as follows. In Chapter 2, we first review CCA and its variants, then specify the population and empirical versions of GAKCCA model and define the contribution of a variable in a group. As the proposed approach requires a regularization parameter, we discuss selection of a regularization parameter. Hypothesis test based on permutation is also introduced. In Chapter 3, we show the results of simulation study to confirm that our method is valid and it explains the relationship of groups well. The discussion is provided in Chapter 4.

Chapter 2

Model

In this chapter, we first briefly review CCA and its variants. Then, we present our GAKCCA method and describe the algorithm for implementation.

2.1 Canonical Correlation Analysis and its variants

Consider two multi-variate groups, $\boldsymbol{X}_1 = (X_{11}, \cdots, X_{1p_1})$ and $\boldsymbol{X}_2 = (X_{21}, \cdots, X_{1p_2})$. $\boldsymbol{X}_1$ has p_1 variables and $\boldsymbol{X}_2$ has p_2 variables. Canonical correlation analysis finds linear combination of each group that maximizes correlation between two linear combinations. That is, CCA finds $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ that optimize the following equation:

$$\max_{\boldsymbol{b}_1, \boldsymbol{b}_2} \mathrm{Cov}\left(\boldsymbol{b}_1^T \boldsymbol{X}_1, \boldsymbol{b}_2^T \boldsymbol{X}_2\right) \tag{2.1}$$

subject to $\mathrm{Var}\left(\boldsymbol{b}_1^T \boldsymbol{X}_1\right) = \mathrm{Var}\left(\boldsymbol{b}_2^T \boldsymbol{X}_2\right) = 1$, where each $\boldsymbol{b}_j \in \mathbb{R}^{p_j}$ for $j = 1, 2$. Variance constraints are due to reduce the freedom of scaling of $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. The above expression is equivalent to maximize $\boldsymbol{c}_1^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \boldsymbol{c}_2$ with constraints $\boldsymbol{c}_1^T \boldsymbol{c}_1 = \boldsymbol{c}_2^T \boldsymbol{c}_2 = 1$ where $\Sigma_{11} = \mathrm{Var}\left(\boldsymbol{X}_1\right)$, $\Sigma_{22} = \mathrm{Var}\left(\boldsymbol{X}_2\right)$ and $\Sigma_{12} = \mathrm{Cov}\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right)$ so that one can solve it by using Cauchy-schwarz inequality to obtain the coefficients. Solutions of (2.1), named $\widehat{\boldsymbol{b}}_1$ and $\widehat{\boldsymbol{b}}_2$, are called (first) canonical coefficients and $\widehat{\boldsymbol{b}}_1^T \boldsymbol{X}_1$ and $\widehat{\boldsymbol{b}}_2^T \boldsymbol{X}_2$ are called (first) canonical variates.

Instead of linear combination restriction in CCA, Kernel canonical correlation analysis utilizes nonlinear functions to extract the relationship between

two groups. KCCA can be formulated as follow:

$$\max_{f_1,f_2} \mathrm{Cov}\Big(f_1(\boldsymbol{X}_1), f_2(\boldsymbol{X}_2)\Big) \tag{2.2}$$

subject to $\mathrm{Var}\Big(f_1(\boldsymbol{X}_1)\Big) = \mathrm{Var}\Big(f_2(\boldsymbol{X}_2)\Big) = 1$, where each $f_j : \mathbb{R}^{p_j} \to \mathbb{R}$ for $j = 1, 2$ is an unknown function in the reproducing kernel Hilbert space (RKHS) ([3]). Given n samples of $\{\boldsymbol{X}_1, \boldsymbol{X}_2\}$, denoted by $\{\boldsymbol{X}_1^{(1)}, \boldsymbol{X}_2^{(1)}\}, \cdots, \{\boldsymbol{X}_1^{(n)}, \boldsymbol{X}_2^{(n)}\}$, f_j can be expressed as $f_j(x) = \sum_{i=1}^{n} a_j^{(i)} \phi_j^{(i)}(\boldsymbol{X}_j^{(i)}, x) + f_j^{perp}(x)$ for $j = 1, 2$ where $\boldsymbol{a}_j = (a_j^{(1)}, \cdots, a_j^{(n)})$ is a coefficient vector, ϕ_j is the kernel corresponding to RKHS that f_j is involved in and f_j^{perp} is orthogonal to linear space spanned by $\phi_j^{(1)}(\boldsymbol{X}_j^{(1)}, \cdot), \cdots, \phi_j^{(n)}(\boldsymbol{X}_j^{(n)}, \cdot)$. By introducing $n \times n$ Gram matrix $\boldsymbol{K}_j$ whose (i, i')-component is $\phi_j(\boldsymbol{X}_j^i, \boldsymbol{X}_j^{i'})$ for $j = 1, 2$ ([17]), empirical version of KCCA can be expressed as

$$\max_{\boldsymbol{a}_1,\boldsymbol{a}_2} \frac{1}{n} \boldsymbol{a}_1^T \boldsymbol{K}_1 \boldsymbol{K}_2 \boldsymbol{a}_2 \tag{2.3}$$

subject to $(1/n)\boldsymbol{a}_1^T \boldsymbol{K}_1 \boldsymbol{K}_1 \boldsymbol{a}_1 = (1/n)\boldsymbol{a}_2^T \boldsymbol{K}_2 \boldsymbol{K}_2 \boldsymbol{a}_2 = 1$. This is similar to CCA optimization problem, so we can find optimal $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ using Cauchy-schwarz inequality idea.

Note that both CCA and KCCA assume two groups of variables. To expand beyond two groups, Kettenring (1971) [10] suggested multi-group generalization of CCA (GCCA or MCCA). GCCA finds linear combinations of each group that optimize certain criterion to reveal multi-group structure. Tenenhaus and Tenenhaus (2011) [21] extended GCCA to a regularization version by imposing constraints on the norm of a coefficient vector in a linear combination as well as the variance of the linear combination (RGCCA). More specifically, given J multi-variate groups $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_J$, RGCCA finds $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_J$ by considering

$$\max_{\boldsymbol{b}_1,\cdots,\boldsymbol{b}_J} \sum_{j,k=1;j\neq k}^{J} c_{jk} g\left[\mathrm{Cov}\Big(\boldsymbol{b}_j^T \boldsymbol{X}_j, \boldsymbol{b}_k^T \boldsymbol{X}_k\Big)\right] \tag{2.4}$$

subject to $\tau_j||\boldsymbol{b}_j||^2+(1-\tau_j)\mathrm{Var}\Big(\boldsymbol{b}_j^T\boldsymbol{X}_j\Big)=1$ for $j=1,\cdots,J$, where g is called a scheme function. Scheme function is related to criterion for selecting canonical variates ([10]). The examples of g are $g(x)=x$ (Horst scheme, [11]), $g(x)=|x|$ (Centroid scheme, [23]) or $g(x)=x^2$ (Factorial scheme, [13]). c_{jk} is an element of $J\times J$ design matrix, C, where $c_{jk}=1$ if j and k groups are related and $c_{jk}=0$, otherwise. $\boldsymbol{\tau}=(\tau_1,\cdots,\tau_J)$ is a regularization parameter (or shrinkage paramter). Regularization parameter enables operation inversion by avoiding ill-conditioned variance matrices ([21, 20]). All τ_j's are between 0 and 1.

Also, Tenenhaus *et al.* (2015) [20] developed a nonlinear version of GCCA (KGCCA) by considering functions of groups. That is, KGCCA finds $f_1,\cdots,f_J$ that optimize

$$\max_{f_1,\cdots,f_J}\sum_{j,k=1;j\neq k}^{J}c_{jk}g\left[\mathrm{Cov}\Big(f_j(\boldsymbol{X}_j),f_k(\boldsymbol{X}_k)\Big)\right] \tag{2.5}$$

subject to $\mathrm{Var}\Big(f_j(\boldsymbol{X}_j)\Big)=1$ for $j=1,\cdots,J$, where each f_j is a real-valued function in reproduced kernel Hilbert space (RKHS). g and c_{jk} are the same as those in RGCCA.

In the next section, we introduce our approach that consider an additive structure in the multi-groups setting.

2.2 Generalized Additive Kernel Canonical Correlation Analysis

As in the previous section, we consider J multivariate random variable groups $\boldsymbol{X}_j=\big(X_{j1},\cdots,X_{jp_j}\big)\in\mathbb{R}^{p_j}$ for $j=1,\cdots,J$. KCCA considers a function on the j-th group variable, $f_j(\boldsymbol{X}_j)$ where f_j is a nonlinear function in RKHS. In our approach, called GAKCCA, we assume that f_j is an additive function in

RKHS as in Balakrishnan *et al.* (2012) [5]. That is,

$$f_j \in \mathcal{H}_j = \left\{ h_j \;\middle|\; h_j(x_1, \cdots, x_{p_j}) = \sum_{l=1}^{p_j} h_{jl}(x_l) \text{ and } h_{jl} \in \mathcal{H}_{jl} \right\} \tag{2.6}$$

where each $\mathcal{H}_{jl}$ is a RHKS with a kernel $\phi_{jl}(\cdot, \cdot)$.

Then, GAKCCA finds $f_j \in \mathcal{H}_j$ that optimize

$$\max_{f_1, \cdots, f_J} \sum_{j,k=1; j \neq k}^{J} c_{jk} g\Big[\mathrm{Cov}\Big(f_j(\boldsymbol{X}_j), f_k(\boldsymbol{X}_k) \Big) \Big] \tag{2.7}$$

subject to $\mathrm{Var}\Big(f_j(\boldsymbol{X}_j) \Big) = 1$ for $j = 1, \cdots, J$, where g and c_{jk} are a scheme function and an element of the design matrix, respectively. Since we assume $f_j \in \mathcal{H}_j$, we can write $f_j(\boldsymbol{X}_j) = \sum_{l=1}^{p_j} f_{jl}(X_{jl})$ so that (2.7) becomes

$$\max_{f_{11}, \cdots, f_{1p_1} \cdots f_{J1} \cdots f_{Jp_J}} \sum_{j,k=1; j \neq k}^{J} c_{jk} g \left[\sum_{l=1}^{p_j} \sum_{m=1}^{p_k} \mathrm{Cov}\Big(f_{jl}(X_{jl}), f_{km}(X_{km}) \Big) \right] \tag{2.8}$$

subject to $\sum_{l=1}^{p_j} \sum_{l'=1}^{p_j} \mathrm{Cov}\Big(f_{jl}(X_{jl}), f_{jl'}(X_{jl'}) \Big) = 1$ for $j = 1, \cdots, J$. We denote $\max_{f_{11}, \cdots, f_{1p_1} \cdots f_{J1} \cdots f_{Jp_J}} \sum_{j,k=1; j \neq k}^{J} c_{jk} g \left[\sum_{l=1}^{p_j} \sum_{m=1}^{p_k} \mathrm{Cov}\Big(f_{jl}(X_{jl}), f_{km}(X_{km}) \Big) \right]$ by $\rho_{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_J}$.

When we introduce a covariance operator on RKHS, mathematical treatment can be simpler ([4, 7, 20]). Let $\Sigma_{jl,km}$ be a covariance operator such that $\mathrm{Cov}\left(f_{jl}(X_{jl}), f_{km}(X_{km}) \right) = \langle f_{jl}, \Sigma_{jl,km} f_{km} \rangle_{\mathcal{H}_{jl}}$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}_{jl}}$ is an inner product on $\mathcal{H}_{jl}$. Then, the equation (2.8) can be expressed as

$$\rho_{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_J} = \max_{f_{11}, \cdots, f_{1p_1} \cdots f_{J1} \cdots f_{Jp_J}} \sum_{j,k=1; j \neq k}^{J} c_{jk} g \left[\sum_{l=1}^{p_j} \sum_{m=1}^{p_k} \langle f_{jl}, \Sigma_{jl,km} f_{km} \rangle_{\mathcal{H}_{jl}} \right] \tag{2.9}$$

subject to $\sum_{l=1}^{p_j} \sum_{l'=1}^{p_j} \langle f_{jl}, \Sigma_{jl,jl'} f_{jl'} \rangle_{\mathcal{H}_{jl}} = 1$ for $j = 1, \cdots, J$.

Note that the equation (2.9) is a theoretical expression. Suppose that we have n samples of $\{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_J\}$ and the i-th sample of $\boldsymbol{X}_j$ is denoted by $\boldsymbol{X}_j^{(i)} = \left(X_{j1}^{(i)}, \cdots, X_{jp_j}^{(i)} \right)$. Fukumizu *et al.* (2007) [6] suggested an estimated

covariance operator $\widehat{\Sigma}_{jl,km}$ that satisfies the following property using kernels ϕ_{jl} and ϕ_{km}.

$$\begin{aligned}\widehat{\mathrm{Cov}}\left(f_{jl}(X_{jl}), f_{km}(X_{km})\right) &= \langle f_{jl}, \widehat{\Sigma}_{jl,km} f_{km}\rangle_{\mathcal{H}_{jl}} \\ &= \frac{1}{n}\sum_{i=1}^{n}\langle f_{jl}, \widehat{\phi}_{jl}^{(i)}\rangle_{\mathcal{H}_{jl}}\langle f_{km}, \widehat{\phi}_{km}^{(i)}\rangle_{\mathcal{H}_{km}} \quad (2.10)\end{aligned}$$

where $\widehat{\phi}_{jl}^{(i)} = \phi_{jl}^{(i)} - \frac{1}{n}\sum_{\xi=1}^{n}\phi_{jl}^{(\xi)}$ and $\phi_{jl}^{(i)} = \phi_{jl}(\cdot, X_{jl}^{(i)})$. Bach and Jordan (2002) [3] then utilized the linear space spanned by $\widehat{\phi}_{jl}^{(1)}, \cdots, \widehat{\phi}_{jl}^{(n)}$ denoted by $\mathcal{S}_{jl}$ to write $f_{jl} = \sum_{i=1}^{n}\alpha_{jl}^{(i)}\widehat{\phi}_{jl}^{(i)} + f_{jl}^{perp}$, where $\alpha_{jl}^{(i)}$ is a coefficient corresponding to $\widehat{\phi}_{jl}^{(i)}$ which needs to be estimated and f_{jl}^{perp} is orthogonal to $\mathcal{S}_{jl}$. Then the equation (2.10) can be expressed as

$$\begin{aligned}&\widehat{\mathrm{Cov}}\left(f_{jl}(X_{jl}), f_{km}(X_{km})\right) \\ =\ & \frac{1}{n}\sum_{i=1}^{n}\left\langle\sum_{i'=1}^{n}\alpha_{jl}^{(i')}\widehat{\phi}_{jl}^{(i')} + f_{jl}^{perp}, \widehat{\phi}_{jl}^{(i)}\right\rangle_{\mathcal{H}_{jl}}\left\langle\sum_{i''=1}^{n}\alpha_{km}^{(i'')}\widehat{\phi}_{km}^{(i')} + f_{km}^{perp}, \widehat{\phi}_{km}^{(i)}\right\rangle_{\mathcal{H}_{km}} \\ =\ & \frac{1}{n}\sum_{i=1}^{n}\left\langle\sum_{i'=1}^{n}\alpha_{jl}^{(i')}\widehat{\phi}_{jl}^{(i')}, \widehat{\phi}_{jl}^{(i)}\right\rangle_{\mathcal{H}_{km}}\left\langle\sum_{i''=1}^{n}\alpha_{km}^{(i'')}\widehat{\phi}_{km}^{(i')}, \widehat{\phi}_{km}^{(i)}\right\rangle_{\mathcal{H}_{km}} \\ =\ & \frac{1}{n}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{i''=1}^{n}\alpha_{jl}^{(i')}\left\langle\widehat{\phi}_{jl}^{(i')}, \widehat{\phi}_{jl}^{(i)}\right\rangle_{\mathcal{H}_{jl}}\left\langle\widehat{\phi}_{km}^{(i'')}, \widehat{\phi}_{km}^{(i)}\right\rangle_{\mathcal{H}_{jl}}\alpha_{km}^{(i'')}. \quad (2.11)\end{aligned}$$

Let us introduce a $n \times n$ symmetric Gram matrix $\boldsymbol{K}_{jl}$ ([17]) whose (i, i')-component is $(\boldsymbol{K}_{jl})_{(i,i')} = \phi_{jl}(X_{jl}^{(i)}, X_{jl}^{(i')}) = \left\langle\phi_{jl}^{(i)}, \phi_{jl}^{(i')}\right\rangle_{\mathcal{H}_{jl}}$. Considering the $n \times n$ identity matrix, $\mathbf{I}_n$, and the n-dimensional vector of ones, $\mathbf{1}_n$, the centered $\boldsymbol{K}_{jl}$, $\widehat{\boldsymbol{K}}_{jl} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)^T \boldsymbol{K}_{jl}\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)$ has the (i, i')-component as follows:

$$
\begin{aligned}
(\widehat{\boldsymbol{K}}_{jl})_{(i,i')} &= \left\langle \phi_{jl}^{(i)}, \phi_{jl}^{(i')} \right\rangle_{\mathcal{H}_{jl}} - \frac{1}{n}\sum_{\xi=1}^{n}\left\langle \phi_{jl}^{(i)}, \phi_{jl}^{(\xi)} \right\rangle_{\mathcal{H}_{jl}} - \frac{1}{n}\sum_{\eta=1}^{n}\left\langle \phi_{jl}^{(\eta)}, \phi_{jl}^{(i')} \right\rangle_{\mathcal{H}_{jl}} \\
&\quad + \frac{1}{n^2}\sum_{\eta=1}^{n}\sum_{\xi=1}^{n}\left\langle \phi_{jl}^{(\eta)}, \phi_{jl}^{(\xi)} \right\rangle_{\mathcal{H}_{jl}} \\
&= \left\langle \phi_{jl}^{(i)}, \phi_{jl}^{(i')} - \frac{1}{n}\sum_{\xi=1}^{n}\phi_{jl}^{(\xi)} \right\rangle_{\mathcal{H}_{jl}} + \left\langle -\frac{1}{n}\sum_{\eta=1}^{n}\phi_{jl}^{(\eta)}, \phi_{jl}^{(i')} - \frac{1}{n}\sum_{\xi=1}^{n}\phi_{jl}^{(\xi)} \right\rangle_{\mathcal{H}_{jl}} \\
&= \left\langle \phi_{jl}^{(i)} - \frac{1}{n}\sum_{\eta=1}^{n}\phi_{jl}^{(\eta)}, \phi_{jl}^{(i')} - \frac{1}{n}\sum_{\xi=1}^{n}\phi_{jl}^{(\xi)} \right\rangle_{\mathcal{H}_{jl}} = \left\langle \widehat{\phi}_{jl}^{(i)}, \widehat{\phi}_{jl}^{(i')} \right\rangle_{\mathcal{H}_{jl}}.
\end{aligned}
$$

Thus, we can further express (2.11) as follows:

$$
\begin{aligned}
\widehat{\mathrm{Cov}}\left(f_{jl}(X_{jl}), f_{km}(X_{km})\right) s &= \frac{1}{n}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{i''=1}^{n}\alpha_{jl}^{(i')}(\widehat{\boldsymbol{K}}_{jl})_{(i',i)}(\widehat{\boldsymbol{K}}_{km})_{(i,i'')}\alpha_{km}^{(i'')} \\
&= \frac{1}{n}\boldsymbol{a}_{jl}^{T}\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{km}\boldsymbol{a}_{km},
\end{aligned}
$$

where $\boldsymbol{a}_{jl} = \left(a_{jl}^{(1)}, \cdots, a_{jl}^{(n)}\right)^{T}$.

Note that the centered Gram matrix $\widehat{\boldsymbol{K}}_{jl}$ is singular since the sum of rows or columns is zero. Thus, the constraint $\sum_{l=1}^{p_j}\sum_{l'=1}^{p_j}\left\langle f_{jl}, \widehat{\Sigma}_{jl,jl'} f_{jl'}\right\rangle = \sum_{l=1}^{p_j}\sum_{l'=1}^{p_j}\frac{1}{n}\boldsymbol{a}_{jl}^{T}\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{jl'}\boldsymbol{a}_{jl'} = 1$ does not provide a unique solution to our method. So, similar to KCCA methods ([3], [20]), we introduce regularization parameters $\tau_j > 0$ in the constraint conditions such that

$$
\sum_{l=1}^{p_j}\sum_{l'=1}^{p_j}\left\langle f_{jl}, \left\{(1-\tau_j)\widehat{\Sigma}_{jl,jl'} + \tau_j \mathbf{I}_{jl,jl'}\right\} f_{jl'}\right\rangle_{\mathcal{H}_{jl}} = 1, \qquad j = 1, \cdots, J, \quad (2.12)
$$

where $\mathbf{I}_{jl,jl'}$ is an identity operator if $l = l'$ and a zero operator, otherwise.

With the $\widehat{\boldsymbol{K}}_{jl}$, (2.12) can be rewritten as

$$
(1-\tau_j)\sum_{l=1}^{p_j}\sum_{l'=1}^{p_j}\frac{1}{n}\boldsymbol{a}_{jl}^{T}\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{jl'}\boldsymbol{a}_{jl'} + \tau_j\sum_{l=1}^{p_j}\boldsymbol{a}_{jl}^{T}\widehat{\boldsymbol{K}}_{jl}\boldsymbol{a}_{jl} = 1, \qquad j = 1, \cdots, J. \quad (2.13)
$$

In summary, empirical version of GAKCCA (2.9) using the kernel method can be expressed as

$$\widehat{\rho}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J} = \max_{\boldsymbol{a}_{11},\cdots,\boldsymbol{a}_{1p_1}\cdots\boldsymbol{a}_{J1}\cdots\boldsymbol{a}_{Jp_J}} \sum_{j,k=1;j\neq k}^{J} c_{jk} g\left[\sum_{l=1}^{p_j}\sum_{m=1}^{p_k}\frac{1}{n}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{km}\boldsymbol{a}_{km}\right] \tag{2.14}$$

subject to $(1-\tau_j)\sum_{l=1}^{p_j}\sum_{l'=1}^{p_j}\frac{1}{n}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{jl'}\boldsymbol{a}_{jl'} + \tau_j\sum_{l=1}^{p_j}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\boldsymbol{a}_{jl} = 1$, for $j = 1,\cdots,J$.

To find the solution, $\{\widehat{\boldsymbol{a}}_{11},\cdots,\widehat{\boldsymbol{a}}_{1p_1},\cdots,\widehat{\boldsymbol{a}}_{J1},\cdots,\widehat{\boldsymbol{a}}_{Jp_J}\}$ of (2.14), an algorithm similar to the algorithm considered in [20] is applied. Specifically, we introduce a Lagrangian form for the optimization problem (2.14).

$$\begin{aligned}\mathcal{L} &= \sum_{j,k=1;j\neq k}^{J} c_{jk} g\left[\sum_{l=1}^{p_j}\sum_{m=1}^{p_k}\frac{1}{n}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{km}\boldsymbol{a}_{km}\right] \\ &- \sum_{j=1}^{J}\lambda_j\left[(1-\tau_j)\sum_{l=1}^{p_j}\sum_{l'=1}^{p_j}\frac{1}{n}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{jl}\boldsymbol{a}_{jl} + \tau_j\sum_{l=1}^{p_j}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\boldsymbol{a}_{jl} - 1\right],\end{aligned}$$

where λ_j's are Lagrange multipliers. $\frac{\partial\mathcal{L}}{\partial\boldsymbol{a}_{jl'}} = 0$ results in

$$\begin{aligned}&\sum_{k=1;j\neq k}^{J} c_{jk} w\left[\sum_{l=1}^{p_j}\sum_{m=1}^{p_k}\frac{1}{n}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{km}\boldsymbol{a}_{km}\right]\cdot\frac{1}{n}\left(\sum_{m=1}^{p_k}\widehat{\boldsymbol{K}}_{jl'}\widehat{\boldsymbol{K}}_{km}\boldsymbol{a}_{km}\right) \\ =\ &\lambda_j\left\{(1-\tau_j)\frac{1}{n}\sum_{l=1}^{p_j}\widehat{\boldsymbol{K}}_{jl'}\widehat{\boldsymbol{K}}_{jl}\boldsymbol{a}_{jl} + \tau_j\widehat{\boldsymbol{K}}_{jl'}\boldsymbol{a}_{jl'}\right\}.\end{aligned} \tag{2.15}$$

By defining $\boldsymbol{u}_j$ and $\boldsymbol{z}_j$ such as

$$\begin{aligned}\boldsymbol{u}_j &= \sum_{l=1}^{p_j}\widehat{\boldsymbol{K}}_{jl}\boldsymbol{a}_{jl} \\ \boldsymbol{z}_j &= \sum_{k=1;j\neq k}^{J} c_{jk} w\left(\sum_{l=1}^{p_j}\sum_{m=1}^{p_k}\frac{1}{n}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{km}\boldsymbol{a}_{km}\right)\frac{1}{n}\left(\sum_{m=1}^{p_k}\widehat{\boldsymbol{K}}_{km}\boldsymbol{a}_{km}\right) \\ &= \sum_{k=1;j\neq k}^{J} c_{jk} w\left(\sum_{l=1}^{p_j}\frac{1}{n}\boldsymbol{a}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\boldsymbol{u}_k\right)\frac{1}{n}\boldsymbol{u}_k,\end{aligned}$$

The equation (2.15) becomes

$$\widehat{\boldsymbol{K}}_{jl'}\boldsymbol{z}_j = \lambda_j \left[(1-\tau_j)\frac{1}{n}\widehat{\boldsymbol{K}}_{jl'}\boldsymbol{u}_j + \tau_j \widehat{\boldsymbol{K}}_{jl'}\boldsymbol{a}_{jl'} \right]. \tag{2.16}$$

Then, one of solutions of (2.16) is

$$\boldsymbol{a}_{jl'} = \frac{1}{\tau_j}\frac{1}{\lambda_j}\boldsymbol{z}_j - \frac{1}{n}\frac{1-\tau_j}{\tau_j}\boldsymbol{u}_j. \tag{2.17}$$

Next, using $\boldsymbol{u}_j$, the constraint conditions (2.14) becomes

$$(1-\tau_j)\frac{1}{n}\sum_{l=1}^{p_j} \boldsymbol{a}_{jl}^T \widehat{\boldsymbol{K}}_{jl}\boldsymbol{u}_j + \tau_j \sum_{l=1}^{p_j} \boldsymbol{a}_{jl}^T \widehat{\boldsymbol{K}}_{jl}\boldsymbol{a}_{jl} = 1, \quad j = 1, \cdots, J.$$

By letting $r_j = \sum_{l=1}^{p_j} \boldsymbol{z}_j^T \widehat{\boldsymbol{K}}_{jl}\boldsymbol{z}_j$, $s_j = \sum_{l=1}^{p_j} \boldsymbol{u}_j^T \widehat{\boldsymbol{K}}_{jl}\boldsymbol{z}_j$ and $t_j = \sum_{l=1}^{p_j} \boldsymbol{u}_j^T \widehat{\boldsymbol{K}}_{jl}\boldsymbol{u}_j$, the above equation is converted to

$$\begin{aligned}(1-\tau_j)\frac{1}{n}\sum_{l=1}^{p_j} &\left[\frac{1}{\tau_j}\frac{1}{\lambda_j}\boldsymbol{z}_j - \frac{1}{n}\frac{1-\tau_j}{\tau_j}\boldsymbol{u}_j \right]^T \widehat{\boldsymbol{K}}_{jl}\boldsymbol{u}_j \\ &+ \tau_j \sum_{l=1}^{p_j} \left[\frac{1}{\tau_j}\frac{1}{\lambda_j}\boldsymbol{z}_j - \frac{1}{n}\frac{1-\tau_j}{\tau_j}\boldsymbol{u}_j \right]^T \widehat{\boldsymbol{K}}_{jl} \left[\frac{1}{\tau_j}\frac{1}{\lambda_j}\boldsymbol{z}_j - \frac{1}{n}\frac{1-\tau_j}{\tau_j}\boldsymbol{u}_j \right] = 1,\end{aligned}$$

which is equivalent to

$$\frac{1}{\lambda_j}\frac{1-\tau_j}{\tau_j}\frac{1}{n}s_j - \frac{(1-\tau_j)^2}{\tau_j}\frac{1}{n^2}t_j + \frac{1}{\lambda_j^2}\frac{1}{\tau_j}r_j - \frac{2}{\lambda_j}\frac{1-\tau_j}{\tau_j}\frac{1}{n}s_j + \frac{(1-\tau_j)^2}{\tau_j}\frac{1}{n^2}t_j = 1$$

and

$$\lambda_j^2 + \frac{1-\tau_j}{\tau_j}\frac{1}{n}s_j\lambda_j - \frac{1}{\tau_j}r_j = 0. \tag{2.18}$$

Note that (2.18) is a quadratic equation with respect to λ_j, so λ_j can be expressed as the function of τ_j, r_j and s_j. Thus, given $\boldsymbol{a}_{11}, \cdots, \boldsymbol{a}_{1p_1}, \cdots, \boldsymbol{a}_{J1} \cdots \boldsymbol{a}_{Jp_J}$ which satisfy the constraint condition (2.14), we can calculate $\lambda_1, \cdots, \lambda_J$ using the equation (2.18) and in return, we calculate $\boldsymbol{a}_{11}, \cdots, \boldsymbol{a}_{1p_1}, \cdots, \boldsymbol{a}_{J1} \cdots \boldsymbol{a}_{Jp_J}$ from the equation (2.17). This recursive procedure stops when a convergence

criterion is satisfied and we obtain $\widehat{\rho}_{\boldsymbol{X}_1,\cdots\boldsymbol{X}_J}$ and $\widehat{\boldsymbol{a}}_{11},\cdots,\widehat{\boldsymbol{a}}_{1p_1},\cdots,\widehat{\boldsymbol{a}}_{J1}\cdots\widehat{\boldsymbol{a}}_{Jp_J}$. The detailed algorithm is described in the Appendix.

In classical CCA, the contribution of a variable in a group in relation between the group and the other group is measured by correlation ([18]). To be specific, we consider the contribution of X_{1l} in $\boldsymbol{X}_1$ for the relation between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ as $\mathrm{Corr}(\widehat{b}_{1l}X_{1l},\widehat{\boldsymbol{b}}_2^T\boldsymbol{X}_2)$, where $\widehat{\boldsymbol{b}}_1=(\widehat{b}_{11},\cdots,\widehat{b}_{1p_1})$ and $\widehat{\boldsymbol{b}}_2$ are canonical coefficients. A high absolute value of $\mathrm{Corr}(\widehat{b}_{1l}X_{1l},\widehat{\boldsymbol{b}}_2^T\boldsymbol{X}_2)$ implies that X_{1l} plays a significant role in the relation between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

Similarly, we can measure the contribution of a variable in a group in relation between the group and the other group in our approach, GAKCCA. We define the contribution coefficient of the lth variable in the jth group, X_{jl} in relation between $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$ as

$$r_{X_{jl},\boldsymbol{X}_k} \;=\; \mathrm{Corr}\Big(f_{jl}(X_{jl}),f_k(\boldsymbol{X}_k)\Big).$$

We also define the measure for the relation between $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$ as

$$r_{\boldsymbol{X}_j,\boldsymbol{X}_k} \;=\; \mathrm{Corr}\Big(f_j(\boldsymbol{X}_j),f_k(\boldsymbol{X}_k)\Big).$$

The empirical version of $r_{X_{jl},\boldsymbol{X}_k}$ and $r_{\boldsymbol{X},\boldsymbol{X}_k}$ can be formulated as

$$\begin{aligned}
\widehat{r}_{X_{jl},\boldsymbol{X}_k} &= \widehat{\mathrm{Corr}}\Big(f_{jl}(X_{jl}),f_k(\boldsymbol{X}_k)\Big)\\
&= \frac{\widehat{\mathrm{Cov}}\Big(f_{jl}(X_{jl}),f_k(\boldsymbol{X}_k)\Big)}{\sqrt{\widehat{\mathrm{Var}}\Big(f_{jl}(X_{jl})\Big)}\sqrt{\widehat{\mathrm{Var}}\Big(f_k(\boldsymbol{X}_k)\Big)}}\\
&= \frac{\displaystyle\sum_{m=1}^{p_k}\widehat{\boldsymbol{a}}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{km}\widehat{\boldsymbol{a}}_{km}}{\sqrt{\widehat{\boldsymbol{a}}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{a}}_{jl}}\sqrt{\displaystyle\sum_{m=1}^{p_k}\sum_{m'=1}^{p_k}\widehat{\boldsymbol{a}}_{km}^T\widehat{\boldsymbol{K}}_{km}\widehat{\boldsymbol{K}}_{km'}\widehat{\boldsymbol{a}}_{km'}}},
\end{aligned}$$

and

$$\widehat{r}_{\boldsymbol{X}_j,\boldsymbol{X}_k} = \frac{\sum_{l=1}^{p_j}\sum_{m=1}^{p_k}\widehat{\boldsymbol{a}}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{km}\widehat{\boldsymbol{a}}_{km}}{\sqrt{\sum_{l=1}^{p_j}\sum_{l'=1}^{p_j}\widehat{\boldsymbol{a}}_{jl}^T\widehat{\boldsymbol{K}}_{jl}\widehat{\boldsymbol{K}}_{jl'}\widehat{\boldsymbol{a}}_{jl'}}\sqrt{\sum_{m=1}^{p_k}\sum_{m'=1}^{p_k}\widehat{\boldsymbol{a}}_{km}^T\widehat{\boldsymbol{K}}_{km}\widehat{\boldsymbol{K}}_{km'}\widehat{\boldsymbol{a}}_{km'}}}.$$

Simulation study shows that empirical contribution coefficient and measure for the relation between two groups describe structural information of variable groups well.

2.3 Regularization Parameter Selection

Choosing appropriate regularization parameters is one of challenges. We consider cross validation idea for selecting regularization parameters for our GAKCCA. Using the whole data, we approximate f_j and denote as $\hat{f}_j$. Using the split data, we approximate f_j and denote as $\hat{f}_j^{-g}$ which is obtained by excluding the gth split. Then, compare these two estimate to select the regularization parameters.

In detail, we describe selection procedure as follows. We split the n samples of $\{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J\}$ into G subsets, denoting $X[1],\cdots,X[G]$, where each $X[g]$ contains n_g samples of $\{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J\}$ and $n_1+\cdots n_G=n$. For each $j=1,\cdots,J$, we estimate f_j by $\hat{f}_j$ and $\hat{f}_j^{-g}$.

$$\begin{aligned}\widehat{f_j} &= \sum_{l=1}^{p_j}\sum_{i=1}^{n}\widehat{a}_{jl}^{(i)}\phi_{jl}^{(i)},\\ \widehat{f_j^{-g}} &= \sum_{l=1}^{p_j}\sum_{i:X_{jl}^{(i)}\notin X[(g)]}\widehat{a}_{jl}^{(i),(-g)}\phi_{jl}^{(i),(-g)},\end{aligned}$$

where $\widehat{a}_{jl}^{(i),(-g)}$ and $\phi_{jl}^{(i),(-g)}$ are calculated from the data excluding $X[g]$ in the

entire dataset. Then, we obtain

$$L(\boldsymbol{\tau}) = L(\tau_1, \cdots, \tau_J) = \frac{1}{G}\sum_{g=1}^{G}\sum_{j=1}^{J}\sum_{x\in X([G])}\left(\frac{\widehat{f}_j(x) - \widehat{f}_j^{(-g)}(x)}{\widehat{f}_j(x)}\right)^2.$$

and selection of τ is made by minimizing $L(\boldsymbol{\tau})$.

The main idea of this procedure is that f_{jl} can be expressed as $f_{jl} = \sum_{i=1}^{n} a_{jl}^{(i)}\widehat{\phi}_{jl}^{(i)} + f_{jl}^{perp}$ by reproducing property in RKHS and we consider $\sum_{i=1}^{n} a_{jl}^{(i)}\widehat{\phi}_{jl}^{(i)}$ as an approximation of f_{jl}. Then cross validation procedure chooses $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_J)$ which τ_j's may not be equal, but for the purpose of simplicity in computation, we assume all τ_j's are equal.

2.4 Permutation test

In classical CCA, Wilks' lambda statistic is widely used ([1]) to test the hypothesis that there is no relationship between two groups. However, multivariate normal distribution assumption of the Wilks' lambda test is difficult to apply to GAKCCA since to test the hypothesis, ordinary CCA has only to check covariance between two group. However, GAKCCA considers covariances of all pair of groups simultaneously. Moreover, nonlinear extension using kernel method makes data structure more complicated. For those reasons, formulating test statistics is rather complex. So, we consider permutation test approach to test $\rho_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J} = 0$. That is, we approximate the sampling distribution of test statistics, $\widehat{\rho}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}$, by obtaining test statistics from resampling under the null hypothesis.

First, from the original data, we calculate $\widehat{\rho}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}$, denoted as $\widehat{\rho}^{obs}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}$. Second, for the j-th group, we sample $\left\{\boldsymbol{X}_j^{(1)*}, \cdots, \boldsymbol{X}_j^{(n)*}\right\}$ from $\left\{\boldsymbol{X}_j^{(1)}, \cdots, \boldsymbol{X}_j^{(n)}\right\}$ with replacement. We do the same procedure for all groups. Note that this resampled $\left\{\boldsymbol{X}_1^{(k)*}, \cdots, \boldsymbol{X}_J^{(k)*}\right\}$ do not necessary keep the order as it should not

matter under the null hypothesis. Third, from the resampled data, we calculate $\widehat{\rho}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}$. Fourth, we repeat second and third steps m times and obtain $\widehat{\rho}^{\{1\}}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}\cdots,\widehat{\rho}^{\{m\}}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}$. Last, we find an empirical distribution $\widehat{F}$ from $\widehat{\rho}^{\{1\}}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}\cdots,\widehat{\rho}^{\{m\}}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}$ and reject the null hypothesis if $1-\widehat{F}\left(\widehat{\rho}^{obs}_{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_J}\right)$ is less than the pre-specified significant level. In this paper, we set $m=300$.

Analogous hypothesis test methods can be applied to test whether a certain variable is helpful for relationship within groups or not via the contribution coefficient.

Chapter 3

Empirical Study

To check the effectiveness of our method, we consider two synthesized data; one is an inter-independent case (Case I) and the other is an inter-dependent case (Case II).

For Case I, we consider 3 blocks of variables ($\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$). The number of members in each block and their distribution assumption are as follows:

- $\boldsymbol{X}_1 = (X_{11}, X_{12}) : X_{11} \sim N(0,1),\ X_{12} \sim N(0,1)$
- $\boldsymbol{X}_2 = (X_{21}, X_{22}, X_{23}, X_{24}) : X_{21} \sim N(0,1), X_{22} \sim N(0,1), X_{23} \sim N(0,1),$ $X_{24} \sim N(0,1)$
- $\boldsymbol{X}_3 = (X_{31}, X_{32}, X_{33}) : X_{31} \sim N(0,1),\ X_{32} \sim N(0,1),\ X_{33} \sim N(0,1)$

Here we assume that all $N(0,1)$s are independent so that 3 blocks $\boldsymbol{X}_1, \boldsymbol{X}_2$ and $\boldsymbol{X}_3$ are mutually independent. From this setting, we generate 100 data points, that is, the number of samples is 100 ($n = 100$).

To apply our method, GAKCCA, we use a Gaussian kernel for each variable. A Gaussian kernel for the lth variable in the jth block is given as $\phi_{jl}(x, y) = -\frac{||x-y||^2}{2\sigma_{jl}^2}$, where σ_{jl} can be viewed as a bandwidth. We set σ_{jl} by median distance between data points in $\{X_{jl}^{(1)}, \cdots, X_{jl}^{(n)}\}$ as in Balakrishnan *et al.* (2012) [5] or Tenenhaus *et al.* (2015) [20]. Also, we use a fully-connected design matrix, that is, $c_{jk} = 1$ if $j \neq k$ and $c_{jk} = 0$,otherwise. Then, we adopt a Horst scheme

function, $g(x) = x$. Without further notice, Gaussian kernel with median-based bandwidth, fully-connected design matrix and Horst scheme function will be used in all simulation study and real data application in this paper.

For the simulated data, we obtain estimates of $\rho_{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3}$, $r_{\boldsymbol{X}_1,\boldsymbol{X}_2}$, $r_{\boldsymbol{X}_2,\boldsymbol{X}_3}$ and $r_{\boldsymbol{X}_3,\boldsymbol{X}_1}$. Then by the permutation test described in the previous section, we calculate a p-value for each estimate. We repeat this procedure 300 times using 300 simulated data. The results are given in Table 3.1. The Table shows that there is no significant relationship between 3 blocks (p-value of $\widehat{\rho}_{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3}$ is 0.517 in average), which correctly captures dependence/independence of the simulation setting for Case I.

Table 3.1: Averages of estimated values and the corresponding p-values from the permutation test over 100 simulated data for Case I (Independent case). The number in parentheses is standard deviation over 100 simulated data.

	Estimate	p-value
$\rho_{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3}$	0.544 (0.210)	0.517 (0.268)
$r_{\boldsymbol{X}_1,\boldsymbol{X}_2}$	0.303 (0.115)	0.447 (0.289)
$r_{\boldsymbol{X}_2,\boldsymbol{X}_3}$	0.341 (0.077)	0.465 (0.300)
$r_{\boldsymbol{X}_3,\boldsymbol{X}_1}$	0.287 (0.098)	0.421 (0.286)

In Case II, we consider 3 blocks ($\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, $\boldsymbol{Y}_3$) again. The number of members in each block and their distribution assumption are as follows.

- $\boldsymbol{Y}_1 = (Y_{11}, Y_{12}) : Y_{11} \sim z + N(0,1)$, $Y_{12} \sim N(0,1)$
- $\boldsymbol{Y}_2 = (Y_{21}, Y_{22}, Y_{23}, Y_{24}) : Y_{21} \sim N(0,1)$, $Y_{22} \sim z^2 + N(0,1)$, $Y_{23} \sim N(0,1)$, , $Y_{24} \sim N(0,1)$

- $\boldsymbol{Y}_3 = (Y_{31}, Y_{32}, Y_{33}) : Y_{31} \sim |z| + N(0,1)$, $Y_{32} \sim z\sin(z) + N(0,1)$, $Y_{33} \sim N(0,1)$,

where $z \sim$ uniform$[-5, 5]$. Here we assume all $N(0,1)$s are independent. Since Y_{11}, Y_{22}, Y_{31} and Y_{32} contain functions of z, these four variables are linked with nonlinear relationship.

From this setting, we generate 100 data points, that is, the number of samples is 100 ($n = 100$) and apply our method to estimates of $\rho_{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\boldsymbol{Y}_3}$, $r_{\boldsymbol{Y}_1,\boldsymbol{Y}_2}$, $r_{\boldsymbol{Y}_2,\boldsymbol{Y}_3}$ and $r_{\boldsymbol{Y}_3,\boldsymbol{Y}_1}$ and the corresponding p-values by the permutation test. With 300 simulated data sets, we repeat this procedure. The averages of estimated values and the corresponding p-values are provided in Table 3.2. A small p-value for testing $\rho_{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\boldsymbol{Y}_3} = 0$ indicates that the blocks are related. We can also see from small p-values of $r_{\boldsymbol{Y}_1,\boldsymbol{Y}_2}$, $r_{\boldsymbol{Y}_2,\boldsymbol{Y}_3}$ and $r_{\boldsymbol{Y}_3,\boldsymbol{Y}_1}$, all three groups are inter-related, which capture dependence between blocks correctly for Case II. Note that the value of $\rho_{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\boldsymbol{Y}_3}$ can be larger than one as it is a combination of functions of covariances. On the other hand, the relation measure $r_{\boldsymbol{Y}_j,\boldsymbol{Y}_k}$ should be less than equal to one as it is a correlation.

Table 3.2: Averages of estimated values and the corresponding p-values from the permutation test over 300 simulated data for Case II (dependent case). The number in parentheses is standard deviation over 300 simulated data.

	Estimate	p-value
$\rho_{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\boldsymbol{Y}_3}$	1.992 (0.419)	0.000 (0.001)
$r_{\boldsymbol{Y}_1,\boldsymbol{Y}_2}$	0.779 (0.044)	0.000 (0.000)
$r_{\boldsymbol{Y}_2,\boldsymbol{Y}_3}$	0.911 (0.022)	0.000 (0.000)
$r_{\boldsymbol{Y}_3,\boldsymbol{Y}_1}$	0.728 (0.051)	0.000 (0.000)

To investigate which variables in the block contribute to the relationship, we calculate contribution coefficients, $r_{Y_{jl},\boldsymbol{Y}_k}$ introduced in the previous section. The results are given in Table 3.3. Recall that Y_{11}, Y_{22}, Y_{31} and Y_{32} have a common component z in the simulation setting. We indicate this by a bold font in the first column of Table 3.3. Y_{11} in the first block $\boldsymbol{Y}_1$ is the one that contribute to the relation between $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, and $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_3$. The empirical contribution coefficients and the corresponding p-values show that Y_{11} is contributing to that relationship compared to Y_{12}. Similarly, we can see from Table 3.3 that the empirical contribution coefficients successfully capture the contribution of Y_{22}, Y_{31} and Y_{32} in relation between their corresponding block and the other blocks.

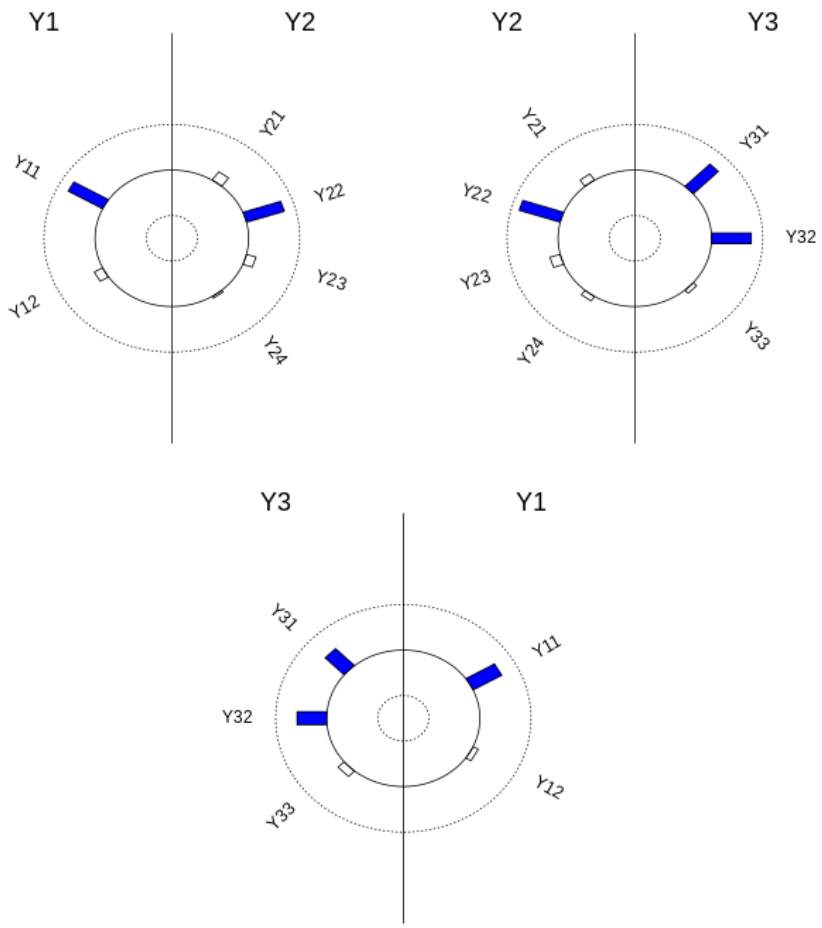
Table 3.3: Averages of empirical contribution coefficients and the corresponding p-values from the permutation test over 100 simulated data for Case II (dependent case).

Estimate / p-value	$\boldsymbol{Y}_1$	$\boldsymbol{Y}_2$	$\boldsymbol{Y}_3$
$\boldsymbol{Y}_{11}$		**0.785/0.000**	**0.730/0.000**
Y_{12}		0.153/0.626	0.155/0.559
Y_{21}	0.146/0.487		0.152/0.540
$\boldsymbol{Y}_{22}$	**0.778/0.000**		**0.915/0.000**
Y_{23}	0.145/0.491		0.150/0.551
Y_{24}	0.145/0.497		0.154/0.511
$\boldsymbol{Y}_{31}$	**0.661/0.000**	**0.788/0.000**	
$\boldsymbol{Y}_{32}$	**0.646/0.000**	**0.849/0.000**	
Y_{33}	0.151/0.500	0.153/0.643	

To visualize contribution of each variable in relation with the other block, we utilize a helio plot. Figure 3.1 is helio plots between pairs of blocks in the second simulation setting (Case II). In the helio plot, variables in two blocks are listed in a circular layout. The size of a bar indicates the value of empirical contribution coefficient of that variable to the other block. For example, in the upper left helio plot in Figure 3.1, the size of the bar corresponding to Y_{11} represents the value of empirical contribution coefficient of Y_{11} to $\boldsymbol{Y}_2$, $\widehat{r}_{Y_{11},\boldsymbol{Y}_2}$. Also, blue colored bars means the p-value of the corresponding empirical contribution coefficient is below 0.05. The upper left helio plot in Figure 3.1 shows that Y_{11} has a significant influence on the relation to $\boldsymbol{Y}_2$ and Y_{12} is less relevance in the relation to $\boldsymbol{Y}_2$. Similarly, from the same helio plot, Y_{22} has a significant influence on the relation to $\boldsymbol{Y}_1$ and the other variables in $\boldsymbol{Y}_2$ except for Y_{22} are less relevance in relation to $\boldsymbol{Y}_1$. From Figure 3.1, we can see that GAKCCA reveal nonlinear relation between blocks and contribution in Case II, properly.

We applied RGCCA to the Case II simulated data (dependent case) for comparison with GAKCCA. We utilized `RGCCA` package ([19]) in **R** (`www.r-project.org`) and implemented the permutation test to extract significant groups. The design matrix, scheme function, number of samples for the permutation test and number of simulated data sets are the same as the ones that we considered for GAKCCA. The RGCCA result shows that there is a significant relationship between Y_2 and Y_3 (The average of empirical correlation between first canonical variate of Y_2 and that of Y_3 is 0.875 with p-value 0.000), but more weak relationship between Y_1 and Y_2, Y_3 and Y_1 than that from GAKCCA (The average of empirical correlations from RGCCA are 0.164, 0.164 with p-value 0.520, 0.577, respectively). The limitation of RGCCA that can only consider linear relationship between groups leads to fail to find clear nonlinear relationship within them.

Figure 3.1: Helio plots of contribution coefficients $r_{Y_{jl},\boldsymbol{Y}_k}$ in Case II. The size of a bar indicates the value of empirical contribution coefficient of that variable to the other block. Blue colored bars means the p-value of the corresponding empirical contribution coefficient is below 0.05.



Chapter 4

Conclusion

In this paper, we have proposed a generalized version of additive kernel CCA. Due to the nature of the objective function, the set of regularization parameters are introduced and we consider the cross validation by comparing estimated additive components for the selection of regularization parameters. A permutation-based test is introduced for checking the relationship between groups. Simulation study shows the proposed method can successfully identify nonlinear relationship between groups and reveals the influence of each variable in the group. Such advantage will be useful in many research areas that deal with multivariate data and in-depth analysis compared to the traditional CCA is possible.

There are some potential extensions of our method. First, classical CCA can consider the second canonical variates that maximizes the correlation $\mathrm{Corr}(b_1^T X_1, b_2^T X_2)$ among all choices that are uncorrelated with the first canonical variates. Introducing the concept of the second canonical variates in GAKCCA could reveal additional structural information within groups that the current GAKCCA model does not explain.

In selecting regularization parameters in GAKCCA, intensive computation is inevitable. Thus, it is worth investigating on developing an algorithm to make computation faster or finding a computationally more efficient selecting

method. Compared with classical CCA, which uses a simple test statistic like Wilk's lambda, permutation test requires more computation time. However, the computation burden can be effectively reduced by distributed computing.

Bibliography

[1] T. W. Anderson. An introduction to multivariate statistical analysis. Technical report, Wiley New York, 1962.

[2] R. Arora and K. Livescu. Kernel cca for multi-view learning of acoustic features using articulatory measurements. In *Symposium on Machine Learning in Speech and Language Processing*, 2012.

[3] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of machine learning research*, 3(Jul):1–48, 2002.

[4] C. R. Baker. Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289, 1973.

[5] S. Balakrishnan, K. Puniyani, and J. Lafferty. Sparse additive functional and kernel cca. *arXiv preprint arXiv:1206.4669*, 2012.

[6] K. Fukumizu, F. R. Bach, and A. Gretton. Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research*, 8(Feb):361–383, 2007.

[7] K. Fukumizu, F. R. Bach, and M. I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces. *Journal of Machine Learning Research*, 5(Jan):73–99, 2004.

[8] H. Hotelling. Relations between two sets of variates. *Biometrika*, 1936.

[9] R. A. Johnson, D. W. Wichern, et al. *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 2002.

[10] J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.

[11] N. Krämer. Analysis of high dimensional data with partial least squares and boosting. 2007.

[12] N. B. Larson, G. D. Jenkins, M. C. Larson, R. A. Vierkant, T. A. Sellers, C. M. Phelan, J. M. Schildkraut, R. Sutphen, P. P. Pharoah, S. A. Gayther, et al. Kernel canonical correlation analysis for assessing gene–gene interactions and application to ovarian cancer. *European Journal of Human Genetics*, 22(1):126, 2014.

[13] J.-B. Lohmöller. *Latent variable path modeling with partial least squares*. Springer Science & Business Media, 2013.

[14] P. S. Moreira, N. C. Santos, N. Sousa, and P. S. Costa. The use of canonical correlation analysis to assess the relationship between executive functioning and verbal memory in older adults. *Gerontology and geriatric medicine*, 1:2333721415602820, 2015.

[15] K. Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[16] F. Sadoughi, H. L. Afshar, A. Olfatbakhsh, and N. Mehrdad. Application of canonical correlation analysis for detecting risk factors leading to recurrence of breast cancer. *Iranian Red Crescent Medical Journal*, 18(3), 2016.

[17] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.

[18] A. Sherry and R. K. Henson. Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of personality assessment*, 84(1):37–48, 2005.

[19] A. Tenenhaus and V. Guillemot. *RGCCA: Regularized and Sparse Generalized Canonical Correlation Analysis for Multiblock Data*, 2017. R package version 2.1.2.

[20] A. Tenenhaus, C. Philippe, and V. Frouin. Kernel generalized canonical correlation analysis. *Computational Statistics & Data Analysis*, 90:114–131, 2015.

[21] A. Tenenhaus and M. Tenenhaus. Regularized generalized canonical correlation analysis. *Psychometrika*, 76(2):257, 2011.

[22] K. A. Tsvetanov, R. N. Henson, L. K. Tyler, A. Razi, L. Geerligs, T. E. Ham, J. B. Rowe, et al. Extrinsic and intrinsic brain network connectivity maintains cognition across the lifespan despite accelerated decay of regional brain activation. *Journal of Neuroscience*, 36(11):3115–3126, 2016.

[23] H. Wold. Partial least squares. s. kotz and nl johnson (eds.), encyclopedia of statistical sciences (vol. 6), 1985.

[24] X. Yang, W. Liu, D. Tao, and J. Cheng. Canonical correlation analysis networks for two-view image recognition. *Information Sciences*, 385:338–352, 2017.

[25] T. Yun and L. Guan. Human emotional state recognition using real 3d visual features from gabor library. *Pattern Recognition*, 46(2):529–538, 2013.

국문초록

다양한 다변량 분석 방법들이 널리 쓰이고 있으며 그 중에 널리 쓰이고 있는 방법론 중에 하나로 정준상관분석 (Canonical correlation analysis, CCA) 이 있다. 정준상관분석은 많은 장점에도 불구하고 선형관계에만 국한되었다는 점, 2개의 그룹에만 적용할 수 있는 점 등의 한계점을 지니고 있다. 이러한 한계를 극복하기 위해 커널 정준상관분석 (Kernel canonical correlation analysis, KCCA), 일반화 정준상관분석 등 여러 변형된 정준상관분석법들이 제안되었다. 본 논문에서는 새로운 모델인 일반화 가법 커널 정준상관분석 (Generalized additive kernel canonical correlation analysis, GAKCCA)를 제안하고자 한다. 비선형 확장을 통한 다중 그룹 사이의 관계를 분석하는 것과 더불어, 일반화 가법 커널 정준상관분석은 각 그룹 내 변수가 그룹간 관계에 어느 정도 기여하는지를 보여줄 수 있으며 이를 통해 심층적인 구조 분석이 가능하다. 또한 시뮬레이션 결과를 통해 일반화 가법 커널 정준상관분석가 다중 그룹들 간의 관계 여부를 나타낼 수 있다는 것을 보여준다.

주요어: 다변량 분석, 일반화 가법 커널 정준상관분석, 다중 그룹 데이터 분석
학번: 2017-25680